Developments in Variational Quality Control

R. James Purser, Xiujuan Su, Runhua Yang, and Yanqiu Zhu
NOAA/NCEP/EMC, College Park, MD 20740-3818, U.S.A. (Email: jim.purser@noaa.gov)

1. Introduction

The underlying principle of variational quality control (VQC) is the realistic assumption that actual errors (including those representation errors attributable to the limitations of resolution) of meteorological measurements used to drive a data assimilation deviate from Gaussianity by having distinctly heavier tails (Purser, 1984; Lorenc and Hammon, 1988; Andersson and Järvinen, 1999). The Bayesian implications for a variational assimilation that seeks to minimize the cost function in the form of the negative log-posterior probability density is that the measurements should be adaptively down-weighted when their departures ('O-A') from the analysis to which they are contributing become relatively large.

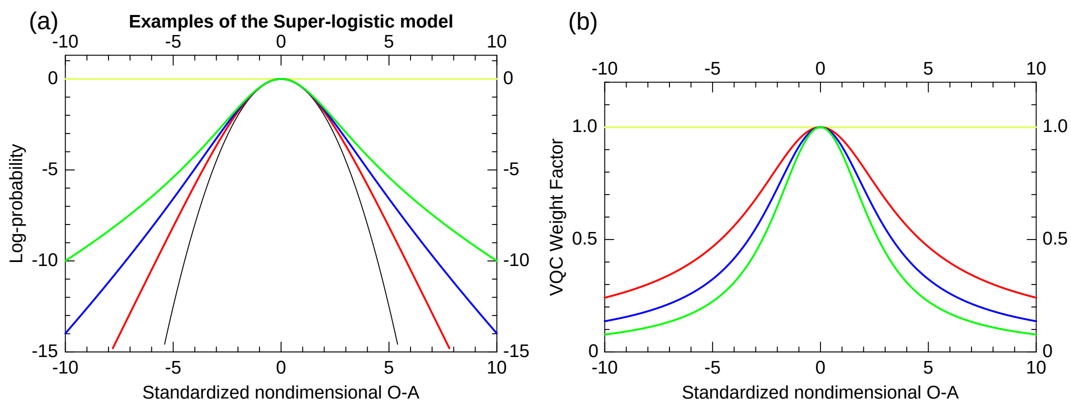

Figure 1. Examples of the standardardized forms of the super-logistic model with neutral tail convexity (red), and increasing amounts of negative convexity (blue and green). Panel (a) shows the log-probability densities, with the Gaussian case included for comparison as the black parabolic curve, while panel (b) shows the profile, also as a function of the standardized O-A, of each corresponding weight factor. The effective weight of a measurement used in the assimilation is the product of this factor and its nominal weight.

2. New developments

We have recently explored the use of a probability model for measurement errors based on a generalization, described in Purser (2011), of the logistic distribution. The classical logistic, or 'sech-squared', density has a convex log-probability profile which resembles the Huber distributions of Tavolato and Isaksen (2014), so that it is incapable of producing multiple minima in the cost function (unlike the constant-plus-Gaussian model of the earliest VQC schemes), and it also seems to fit the error-distributions of many real data types better. However, for some data, the shapes of the tails of the distributions indicate that at least a small degree of concavity is needed, and this is accommodated by the 'super-logistic' generalization of Purser (2011) that we are presently testing within NCEP's Gridpoint Statistical Interpolation. The figure shows examples of the super-logistic model's log-probability with varying degrees of the prescribed convexity (panel a) together with the corresponding effect on the multiplicative factor (panel b) by which the measurements are down-weighted from their standard (Gaussian model) precision weight. The potential risks associated with multiple minima in the cost function will be mitigated in practice by running the first several iterations of the minimization of the cost function using a temporarily neutral convexity parameter, i.e., the ordinary logistic model.

A further development, which the generalization of the logistic model allows us to explore, is the application of these statistical principles to series of coupled measurements. If a set of measurements are assumed to have been made with the same instrument, the detection of a likely gross error in just one of the measurements (an excessively large O-A) can be used to infer a needed down-weighting, not only of itself, but also of its related neighbors even when their own O-A diagnostics alone, are not sufficiently deviant to detect the problem. There are both *in situ* and satellite data types where such an implicit coupling of gross error effects could prove beneficial to the production of more robust and reliable assimilations.

References

Andersson, E., and H. Järvinen 1999 Variational quality control. *Quart. J. Roy. Meteor. Soc.,* **125,** 697–722.

Lorenc, A. C., and O. Hammon 1988 Objective quality control of observations using Bayesian methods. Theory and a practical implementation. *Quart. J. Roy. Meteor. Soc.,* **114,** 515–543.

Purser, R. J. 1984 A new approach to the optimal assimilation of meteorological data by iterative Bayesian analysis. *Preprint, AMS 10th Conference on Weather Forecasting and Analysis,* Clearwater Beach, Florida, pp 102–105.

Purser, R. J. 2011 Mathematical principles of the construction and characterization of a parameterized family of Gaussian mixture distributions suitable to serve as models for the probability distributions of measurement errors in nonlinear quality control. NOAA/NCEP Office Note 468.

Tavolato, C., and L. Isaksen 2014 On the use of a Huber norm for observation quality control in the ECMWF 4D-Var. *Quart. J. Roy. Meteor. Soc.,* **141,** 1514–1527.